

DANIEL J. MULLER, SBN 193396  
dmuller@venturahersey.com  
VENTURA HERSEY & MULLER, LLP  
1506 Hamilton Avenue  
San Jose, California 95125  
Telephone: (408) 512-3022  
Facsimile: (408) 512-3023

*Attorneys for Plaintiff and the Class*

**UNITED STATES DISTRICT COURT**  
**NORTHERN DISTRICT OF CALIFORNIA – SAN FRANCISCO DIVISION**

MICHAEL CHABON, TA-NEHISI COATES, JUNOT DÍAZ, ANDREW SEAN GREER, DAVID HENRY HWANG, MATTHEW Klam, LAURA LIPPMAN, RACHEL LOUISE SNYDER, AYELET WALDMAN, AND JACQUELINE WOODSON.

individually and on behalf of all others similarly situated,

Plaintiffs,

V.

OPENAI, INC., OPENAI, L.P., OPENAI  
OPCO, LLC, OPENAI GP LLP, OPENAI  
STARTUP FUND GP I, LLC, OPENAI  
STARTUP FUND I, LP, and OPENAI  
START UP FUND MANAGEMENT,  
LLC,

## Defendants.

Case No. 3:23-cv-04625 - PHK

# **FIRST AMENDED CLASS ACTION COMPLAINT**

**CLASS ACTION**

## **JURY TRIAL DEMANDED**

1 Plaintiffs Michael Chabon, Ta-Nehisi Coates, Junot Díaz, Andrew Sean Greer, David  
 2 Henry Hwang, Matthew Klam, Laura Lippman, Rachel Louise Snyder, Ayelet Waldman, and  
 3 Jacqueline Woodson (“Plaintiffs”), individually and on behalf of all others similarly situated,  
 4 bring this action against Defendants OpenAI, Inc., OpenAI, LP, OpenAI OpCo, LLC, OpenAI  
 5 GP LLC, OpenAI Startup Fund I, LP, OpenAI Startup Fund GP I, LLC, and OpenAI Startup  
 6 Fund Management, LLC (collectively, “Defendants” or “OpenAI”). Plaintiffs allege as follows  
 7 based upon personal knowledge as to themselves and their own acts, and upon information and  
 8 belief as to all other matters:

9 **NATURE OF ACTION**

10 1. This is a class action lawsuit brought by Plaintiffs on behalf of themselves and a  
 11 Class of authors holding copyrights in their published works arising from OpenAI’s clear  
 12 infringement of their intellectual property.

13 2. OpenAI is a research company specializing in the development of artificial  
 14 intelligence (“AI”) products, such as ChatGPT.

15 3. ChatGPT is an AI chatbot, which produces responses to users’ text queries or  
 16 prompts in a way that mimics human conversation.

17 4. ChatGPT relies on other OpenAI products to function, namely Generative Pre-  
 18 trained Transformer (“GPT”) models. “Generative,” in GPT, represents the model’s ability to  
 19 respond to text inquiries, while “Pre-trained” refers to the model’s use of training datasets to  
 20 program its responses, and “Transformer” concerns the model’s underlying algorithm allowing  
 21 it to function.

22 5. OpenAI has released five versions of GPT models, and the current version of  
 23 ChatGPT runs on GPT-3.5 and GPT-4, depending on whether the user has subscribed to the  
 24 premium version of ChatGPT. Only the version of ChatGPT that runs on GPT-3.5 is available  
 25 at no cost to the public.

26 6. OpenAI’s GPT models are types of “large language model,” which is a form of  
 27 deep-learning algorithm programmed through “training datasets,” consisting of massive  
 28 amounts of text data copied from the internet by OpenAI. The GPT models extract information

from their training datasets in order to learn the statistical relationships between words, phrases, and sentences, which allow them to generate coherent and contextually relevant responses to user prompts or queries.

7. A large language model's responses to user prompts or queries are entirely and uniquely dependent on the text contained in its training dataset, necessarily processing and analyzing the information contained in its training dataset to generate responses.

8. OpenAI incorporated Plaintiffs' and Class members' copyrighted works in datasets used to train its GPT models powering its ChatGPT product. Indeed, when ChatGPT is prompted, it generates not only summaries, but in-depth analyses of the themes present in Plaintiffs' copyrighted works, which is only possible if the underlying GPT model was trained using Plaintiffs' works.

9. Plaintiffs and Class members did not consent to the use of their copyrighted works as training material for GPT models or for use with ChatGPT.

10. Defendants, by and through their operation of ChatGPT, benefit commercially and profit handsomely from their unauthorized and illegal use of Plaintiffs' and Class members' copyrighted works.

## **JURISDICTION AND VENUE**

11. This Court has subject matter jurisdiction of this action pursuant to 28 U.S.C. § 1331 because this case arises under the Copyright Act (17 U.S.C. § 501) and the Digital Millennium Copyright Act (17 U.S.C. § 1202).

12. This Court has personal jurisdiction over Defendants pursuant to 18 U.S.C. §§ 1965(b) & (d), because they maintain their principal places of business in, and are thus residents of, this judicial district, maintain minimum contacts with the United States, this judicial district, and this State, and they intentionally avail themselves of the laws of the United States and this state by conducting a substantial amount of business in California. For these same reasons, venue properly lies in this District pursuant to 28 U.S.C. §§ 1391(a), (b) and (c).

1           **PARTIES**

2           **A. Plaintiffs**

3           13. Plaintiff Michael Chabon (“Plaintiff Chabon”) is a resident of California.  
 4 Plaintiff Chabon is an author who owns registered copyrights in many works, including but not  
 5 limited to, *The Mysteries of Pittsburgh*, *Wonder Boys*, *The Amazing Adventures of Kavalier &*  
 6 *Clay*, *the Yiddish Policemen’s Union*, *Gentlemen of the Road*, *Telegraph Avenue*, *Fight of the*  
 7 *Century*, *Kingdom of Olive and Ash*, and *Moonglow*. Plaintiff Chabon is the recipient of the  
 8 Pulitzer Prize for Fiction, Hugo, Nebula, Los Angeles Times Book Prize, and the National  
 9 Jewish Book Award, among many others achieved over the span of a writing career spanning  
 10 more than 30 years. Plaintiff Chabon’s works include copyright-management information that  
 11 provides information about the copyrighted work, including the title of the work, its ISBN or  
 12 copyright registration number, the name of the author, and the year of publication.

13           14. Plaintiff Ta-Nehisi Coates (“Plaintiff Coates”) is a resident of New York.  
 14 Plaintiff Coates is an author who owns registered copyrights in many works, including but not  
 15 limited to, *Between the World and Me*, *The Beautiful Struggle*, *The Water Dancer*, and *We Were*  
 16 *Eight Years in Power*. Plaintiff Coates’s is the recipient of the National Book Award and a  
 17 MacArthur Fellowship. Plaintiff Coates works include copyright-management information that  
 18 provides information about the copyrighted work, including the title of the work, its ISBN or  
 19 copyright registration number, the name of the author, and the year of publication.

20           15. Plaintiff Junot Díaz (“Plaintiff Diaz”) is a resident of Massachusetts. Plaintiff  
 21 Díaz is an author who owns registered copyrights in many works, including but not limited to,  
 22 *Islandborn*, *The Brief Wondrous Life of Oscar Wao*, *This is How you Lose Her*, and *Drown*.  
 23 Plaintiff Díaz is the recipient of the Pulitzer Prize for Fiction, the National Book Critics Circle  
 24 Award, a MacArthur “Genius” Fellowship, PEN/Malamud Award, Dayton Literary Peace Prize,  
 25 Guggenheim Fellowship, and PEN/O. Henry Award. Plaintiff Díaz is the Nancy Allen Professor  
 26 of Writing at the Massachusetts Institute of Technology. Plaintiff Díaz’s works include  
 27 copyright-management information that provides information about the copyrighted work,  
 28

1 including the title of the work, its ISBN or copyright registration number, the name of the author,  
 2 and the year of publication.

3       16. Andrew Sean Greer (“Plaintiff Greer”) is a resident of California. Plaintiff Greer  
 4 is an author who owns registered copyrights in many works, including but not limited to, *The*  
 5 *Confessions of Max Tivoli*, *How it Was for Me*, *The Story of a Marriage*, *The Impossible Lives*  
 6 *of Greta Wells*, *The Path of Minor Planets*, *Less is Lost*, and *Less, a novel*. Plaintiff Greer is the  
 7 recipient of the Pulitzer Prize for Fiction, a Guggenheim Fellowship, and an NEA grant. Plaintiff  
 8 Greer’s works include copyright-management information that provides information about the  
 9 copyrighted work, including the title of the work, its ISBN or copyright registration number, the  
 10 name of the author, and the year of publication.

11       17. Plaintiff David Henry Hwang (“Plaintiff Hwang”) is a resident of New York.  
 12 Plaintiff Hwang is a playwright and screenwriter who owns registered copyrights in many  
 13 works, including but not limited to, *M. Butterfly*, *Chinglish*, *Yellow Face*, *the Dance and the*  
 14 *Railroad*, and *FOB*, as well as the Broadway musical, *Flower Drum Song* (2002 revival).  
 15 Plaintiff Hwang is a Tony Award winner and three-time nominee, a Grammy Award winner  
 16 who has been twice nominated, a three-time OBIE Award winner, and a three-time finalist for  
 17 the Pulitzer Prize in Drama. Plaintiff Hwang’s works include copyright-management  
 18 information that provides information about the copyrighted work, including the title of the  
 19 work, its ISBN or copyright registration number, the name of the author, and the year of  
 20 publication.

21       18. Plaintiff Matthew Klam (“Plaintiff Klam”) is a resident of Washington D.C.  
 22 Plaintiff Klam is an author who owns registered copyrights in several works, including but not  
 23 limited to, *Who is Rich?*, and *Sam the Cat and Other Stories*. Plaintiff Klam is a recipient of a  
 24 Guggenheim Fellowship, a Robert Bingham/PEN Award, a Whiting Writer’s Award, and a  
 25 National Endowment of the Arts. Plaintiff Klam’s works have been selected as Notable Books  
 26 of the year by *The New York Times*, *The Los Angeles Times*, *the Kansas City Star*, and *the*  
 27 *Washington Post*. Plaintiff Klam’s works include copyright-management information that

28

1 provides information about the copyrighted work, including the title of the work, its ISBN or  
 2 copyright registration number, the name of the author, and the year of publication.

3       19. Plaintiff Laura Lippman (“Plaintiff Lippman”) is a resident of Maryland. Plaintiff  
 4 Lippman is an author who owns registered copyrights in several works, including but not limited  
 5 to, *I'd Know you Anywhere*, *What the Dead Know*, *Hardly Knew Her*, *Another Thing to Fall*,  
 6 *Butchers Hill*, and *Baltimore Blues*. Plaintiff Lippman is a *New York Times* bestseller and the  
 7 recipient of the Edgar Award, the Anthony Award, the Agatha Award, the Quill Award, the  
 8 Shamus Award, and the Nero Wolfe Award. Plaintiff Lippman’s works include copyright-  
 9 management information that provides information about the copyrighted work, including the  
 10 title of the work, its ISBN or copyright registration number, the name of the author, and the year  
 11 of publication.

12       20. Plaintiff Rachel Louise Snyder (“Plaintiff Snyder”) is a resident of Washington,  
 13 D.C. Plaintiff Snyder is an author who owns registered copyrights in many works, including but  
 14 not limited to, *Women We Buried*, *Women We Burned*, *No Visible Bruises – What We Don’t*  
 15 *Know About Domestic Violence Can Kill Us*, *What We’ve Lost is Nothing*, and *Fugitive Denim: A Moving Story of People and Pants in the Borderless World of Global Trade*. Plaintiff Snyder  
 16 is a Guggenheim fellow and the recipient of the J. Anthony Lukas Work-in-Progress Award, the  
 17 Hillman Prize, and the Helen Bernstein Book Award, and was a finalist for the National Book  
 18 Critics Circle Award, *Los Angeles Times* Book Prize, and Kirkus Award. Her work has appeared  
 19 in *The New Yorker*, *The New York Times*, *Slate*, and in many other publications. Plaintiff  
 20 Snyder’s works include copyright-management information that provides information about the  
 21 copyrighted work, including the title of the work, its ISBN or copyright registration number, the  
 22 name of the author, and the year of publication.

24       21. Plaintiff Ayelet Waldman (“Plaintiff Waldman”) is a resident of California.  
 25 Plaintiff Waldman is an author and screen and television writer who owns registered copyrights  
 26 in several works, including but not limited to, *Love and other Impossible Pursuits*, *Red Hook*  
 27 *Road*, *Love and Treasure*, *Bad Mother*, *Daughter’s Keeper*, *A Really Good Day*, *Fight of the*  
 28 *Century*, and *Kingdom of Olives and Ash*. Plaintiff Waldman has been nominated for an Emmy

1 and a Golden Globe and is the recipient of numerous awards including a Peabody, AFI award,  
 2 and a Pen Award, among others. Plaintiff Waldman's works include copyright-management  
 3 information that provides information about the copyrighted work, including the title of the  
 4 work, its ISBN or copyright registration number, the name of the author, and the year of  
 5 publication.

6       22. Plaintiff Jacqueline Woodson ("Plaintiff Woodson") is a resident of New York.  
 7 Plaintiff Woodson is an author who owns registered copyrights in many works, including but  
 8 not limited to, *Brown Girl Dreaming*, *After Tupac & D Foster*, *Harbor Me*, *Before the Ever*  
 9 *After*, *Day You begin*, and *Another Brooklyn*. Plaintiff Woodson is the recipient of the  
 10 MacArthur "Genius" Award, a Guggenheim Fellowship, the National Book Award, a Caldecott  
 11 Honor Medal, the Coretta Scott King Award, the Newbery Honor Medal, and the Hans Christian  
 12 Andersen Award, among others. Plaintiff Woodson's works include copyright-management  
 13 information that provides information about the copyrighted work, including the title of the  
 14 work, its ISBN or copyright registration number, the name of the author, and the year of  
 15 publication.

16       23. At all times relevant hereto, Plaintiffs have been and remain the holders of the  
 17 exclusive rights under the Copyright Act of 1976 (17 U.S.C. §§ 101, *et seq.* and all amendments  
 18 thereto) to reproduce, distribute, display, or license the reproduction, distribution, and/or display  
 19 the works identified in paragraphs 13-22, *supra*.

20 **B. Defendants**

21       24. Defendant OpenAI, Inc. is a Delaware nonprofit corporation with its principal  
 22 place of business located at 3180 18th St., San Francisco, CA 94110.

23       25. Defendant OpenAI, LP is a Delaware limited partnership with its principal place  
 24 of business located at 3180 18th St., San Francisco, CA 94110. OpenAI, LP is a wholly owned  
 25 subsidiary of OpenAI, Inc. that is operated for profit. OpenAI, Inc. controls OpenAI, LP directly  
 26 and through the other OpenAI entities.

27       26. Defendant OpenAI OpCo, LLC is a Delaware limited liability company with its  
 28 principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI

1 OpCo, LLC is a wholly owned subsidiary of OpenAI, Inc. that is operated for profit. OpenAI,  
2 Inc. controls OpenAI OpCo, LLC directly and through the other OpenAI entities.

3 27. Defendant OpenAI GP, LLC is a Delaware limited liability company with its  
4 principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI GP,  
5 LLC is a general partner of OpenAI, LP. OpenAI GP manages and operates the day-to-day  
6 business and affairs of OpenAI, LP. OpenAI GP was aware of the unlawful conduct alleged  
7 herein and exercised control over OpenAI, LP throughout the Class Period. OpenAI, Inc. directly  
8 controls OpenAI GP.

9 28. Defendant OpenAI Startup Fund I, LP is a Delaware limited partnership with its  
10 principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI  
11 Startup Fund I, LP was instrumental in the foundation of OpenAI, LP, including the creation of  
12 its business strategy and providing initial funding. OpenAI Startup Fund I was aware of the  
13 unlawful conduct alleged herein and exercised control over OpenAI, LP throughout the Class  
14 Period.

15 29. Defendant OpenAI Startup Fund GP I, LLC is a Delaware limited liability  
16 company with its principal place of business located at 3180 18th Street, San Francisco, CA  
17 94110. OpenAI Startup Fund GP I, LLC is the general partner of OpenAI Startup Fund I.  
18 OpenAI Startup Fund GP I is a party to the unlawful conduct alleged herein. OpenAI Startup  
19 Fund GP I manages and operates the day-to-day business and affairs of OpenAI Startup Fund I.

20 30. Defendant OpenAI Startup Fund Management, LLC is a Delaware limited  
21 liability company with its principal place of business located at 3180 18th Street, San Francisco,  
22 CA 94110. OpenAI Startup Fund Management, LLC is a party to the unlawful conduct herein.  
23 OpenAI Startup Fund Management was aware of the unlawful conduct alleged herein and  
24 exercised control over OpenAI, LP throughout the Class Period.

25

26

27

28

## **FACTUAL ALLEGATIONS**

## A. OpenAI's Artificial Intelligence Products

31. OpenAI researches, develops, releases, and maintains AI products with the intention that its products “benefit all of humanity.”<sup>1</sup>

32. ChatGPT is among the products OpenAI has developed, engineered, released, and maintained, which utilizes another OpenAI product, GPT models, to respond to text prompts and queries in a natural, coherent, and fluent way through a web interface.

33. OpenAI has released a series of upgrades to its GPT model, including GPT-1 (released June 2018), GPT-2 (February 2019), GPT-3 (May 2020), GPT-3.5 (March 2022), and most recently, GPT-4 (March 2023)<sup>2</sup>.

34. The current version of ChatGPT utilizes both GPT-3.5 and GPT-4; however, the version of ChatGPT that allows users to choose between using GPT-3.5 and GPT-4 is only available to subscribers at a cost of \$20 per month. Otherwise, users are only able to access the version of ChatGPT that relies on the GPT-3.5 model.<sup>3</sup>

35. OpenAI makes ChatGPT available to software developers through an application-programming interface (“API”), which allows developers to write software programs that exchange data with ChatGPT.<sup>4</sup> OpenAI charges developers for access to ChatGPT by the API on the basis of usage.

## B. OpenAI Uses Copyrighted Works in its Training Datasets

36. As mentioned in paragraph 6, *supra*, OpenAI pre-trains its GPT models using a dataset consisting of various sources and content types, including books, plays, articles, and webpage and other written works, to respond accurately to users' prompts and queries.

37. OpenAI has admitted that, of all sources and content types that can be used to train the GPT models, written works, plays and articles are valuable training material because

<sup>1</sup> *About*, OpenAI, <https://openai.com/about>

<sup>2</sup> Fawad Ali, *GPT-1 to GPT-4: Each of OpenAI's GPT Models Explained and Compared*, Make Use Of (Apr. 11, 2023) <https://www.makeuseof.com/gpt-models-explained-and-compared/>

<sup>3</sup> Introducing ChatGPT Plus, OpenAI (Feb. 1, 2023) <https://openai.com/blog/chatgpt-plus>

<sup>4</sup> OpenAI API, OpenAI (June 11, 2020) <https://openai.com/blog/openai-api>

1 they offer the best examples of high-quality, long form writing and “contain[] long stretches of  
 2 contiguous text, which allows the generative model to learn to condition on long-range  
 3 information.”<sup>5</sup>

4       38. Upon information and belief, OpenAI builds the dataset it uses to train its GPT  
 5 models by scraping the internet for text data.

6       39. While casting a wide net across the internet to capture the most comprehensive  
 7 set of content available allows OpenAI to better train its GPT models, this practice necessarily  
 8 leads OpenAI to capture, download, and copy copyrighted written works, plays and articles.

9       40. Among the content OpenAI has scraped from the internet to construct its training  
 10 datasets are Plaintiffs’ copyrighted works.

11       41. In its June 2018 paper introducing the GPT-1 model, *Improving Language*  
 12 *Understanding by Generative Pre-Training*, OpenAI revealed that it trained the GPT-1 model  
 13 using two datasets: “Common Crawl,” which is a massive dataset of web pages containing  
 14 billions of words, and “BookCorpus,” which is a collection of “over 7,000 unique unpublished  
 15 books from a variety of genres including Adventure, Fantasy, and Romance.”<sup>6</sup>

16       42. BookCorpus is a controversial dataset, assembled in 2015 by a team of AI  
 17 researchers funded by Google and Samsung for the sole purpose of training language models  
 18 like GPT by copying written works from a website called Smashwords, which hosts self-  
 19 published novels, making them available to readers at no cost.<sup>7</sup> Despite those novels being  
 20 largely under copyright, they were copied into the BookCorpus dataset without consent, credit,  
 21 or compensation to the authors.<sup>8</sup>

22       43. OpenAI also copied many books while training GPT-3. In the July 2020 paper  
 23 introducing GPT-3, *Language Models are Few-Shot Learners*, OpenAI disclosed, in addition to

---

24       <sup>5</sup> Alec Radford, *Improving Language Understanding by Generative-Pre-Training*, OpenAI  
 25 (June 11, 2018).

26       <sup>6</sup> *Id.*; see also Fawad Ali, *GPT-1 to GPT-4: Each of OpenAI’s GPT Models Explained and*  
 27 *Compared*, Make Use Of (Apr. 11, 2023) <https://www.makeuseof.com/gpt-models-explained-and-compared/>

28       <sup>7</sup> Jack Bandy, *Dirty Secrets of BookCorpus, a Key Dataset in Machine Learning*, Medium  
 (May 12, 2021) <https://towardsdatascience.com/dirty-secrets-of-bookcorpus-a-key-dataset-in-machine-learning-6ee2927e8650>

<sup>8</sup> *Id.*

1 using the “Common Crawl” and “WebText” datasets that capture web pages, 16% of the GPT-  
 2 training dataset came from “two internet-based book corpora,” which OpenAI simply refers  
 3 to as “Books1” and “Books2.”<sup>9</sup>

4 44. OpenAI has never revealed what books are part of the Books1 and Books2  
 5 datasets or how they were obtained. OpenAI has offered a few clues, admitting that these are  
 6 internet-based datasets that are much larger than BookCorpus.<sup>10</sup> Based on the figures provided  
 7 in its GPT-3 introductory paper, Books1 is nine times larger than BookCorpus, meaning it  
 8 contains roughly 63,000 titles, and Books2 is 42 times larger, meaning it contains about 294,000  
 9 titles.<sup>11</sup>

10 45. A limited number of internet-based book corpora exist that contain this much  
 11 material, meaning there are only a handful of possible sources OpenAI could have used to train  
 12 the GPT-3 model.

13 46. Project Gutenberg is an online archive of e-books whose copyrights have expired.  
 14 Project Gutenberg has long been popular for training AI systems due to the lack of copyright. In  
 15 2018, a team of AI researchers created the “Standardized Project Gutenberg Corpus,” which  
 16 contained “more than 50,000 books.”<sup>12</sup> On that information and belief, the OpenAI Books1  
 17 dataset is based on either the Standardized Project Gutenberg Corpus or Project Gutenberg itself,  
 18 because of the roughly similar sizes of the two datasets.

19 47. As for the Books2 dataset, the only “internet-based books corpora” that have ever  
 20 offered that much material are infamous “shadow library” websites, like Library Genesis  
 21 (“LibGen”), Z-Library, Sci-Hub, and Bibliotik, which host massive collections of pirated books,  
 22 research papers, and other text-based materials.<sup>13</sup> The materials aggregated by these websites  
 23 have also been available in bulk through torrent systems.<sup>14</sup>

24 <sup>9</sup> Tom B. Brown, *Language Models are Few-Shot Learners*, OpenAI (July 22, 2020).

25 <sup>10</sup> *Id.* at 9.

26 <sup>11</sup> *Id.*

27 <sup>12</sup> Martin Gerlach, et al., *A standardized Project Gutenberg corpus for statistical analysis of natural language and quantitative linguistics*, Cornell University (Dec. 19, 2018), <https://arxiv.org/pdf/1812.08092.pdf>

28 <sup>13</sup> See Claire Woodcock, ‘Shadow Libraries’ Are Moving Their Pirated Books to The Dark Web After Fed Crackdowns, Vice (Nov. 30, 2022).

<sup>14</sup> *Id.*

1       48. These illegal shadow libraries have long been of interest to the AI-training  
 2 community. For instance, an AI training dataset published in December 2020 by EleutherAI  
 3 called “Books3” includes a recreation of the Bibliotik collection and contains nearly 200,000  
 4 books.<sup>15</sup> On information and belief, the OpenAI Books2 dataset includes books copied from  
 5 these “shadow libraries,” because those are the sources of trainable books most similar in nature  
 6 and size to OpenAI’s description of Books2.

7       49. When OpenAI introduced GPT-4 in March 2023, the introductory paper  
 8 contained no information about the dataset used to train it.<sup>16</sup> Instead, OpenAI claims that,  
 9 “[g]iven both the competitive landscape and the safety implications of large-scale models like  
 10 GPT-4, this report contains no further details about . . . dataset construction.”<sup>17</sup>

11       50. Regarding GPT-4, OpenAI has conceded that it did filter its dataset “to  
 12 specifically reduce the quantity of inappropriate erotic text content,” implying that it again used  
 13 a large dataset containing text works.<sup>18</sup>

14       **C. OpenAI Unlawfully Infringed Plaintiffs’ Copyrights**

15       51. As explained, ChatGPT’s responses to user queries or prompts, like other large  
 16 language models, rely on the data upon which it is trained to generate responsive content. For  
 17 example, if ChatGPT is prompted to generate a writing in the style of a certain author, GPT  
 18 would generate content based on patterns and connections it learned from analysis of that  
 19 author’s work within its training dataset.

20       52. On information and belief, the reason ChatGPT can generate a writing in the style  
 21 of a certain author or accurately summarize a certain copyrighted book and provide in-depth  
 22 analysis of that book is because it was copied by OpenAI and copied and analyzed by the  
 23 underlying GPT model as part of its training data.

24       53. When ChatGPT is prompted to summarize copyrighted written works authored  
 25 by Plaintiffs, it generates accurate, in-depth summaries and analyses of their works.

---

26       <sup>15</sup> See Alex Perry, *A giant online book collection Meta used to train its AI is gone over*  
 27 *copyright issues*, Mashable (Aug. 18, 2023).

28       <sup>16</sup> *GPT-4 Technical Report*, OpenAI (Mar. 27, 2023).

28       <sup>17</sup> *Id.* at 2.

28       <sup>18</sup> *Id.* at 61.

1       54. For example, when prompted, ChatGPT accurately summarized Plaintiff  
 2 Chabon's novel *The Amazing Adventures of Kavalier & Clay*. When prompted to identify  
 3 examples of trauma in the *Amazing Adventures of Kavalier & Clay*, ChatGPT identified six  
 4 specific examples, including how the main character's "experiences in Europe, including  
 5 witnessing the persecution of Jews and the loss of his family, haunt him throughout the story."  
 6 When asked to write a paragraph in the style of *The Amazing Adventures of Kavalier & Clay*,  
 7 ChatGPT generated a passage imitating Plaintiff Chabon's writing style including references to  
 8 the characters dealing with "the weight of the world at war." *Exhibit A*.

9       55. Likewise, ChatGPT accurately summarized Plaintiff Coates novel *The Water*  
 10 *Dancer*. When prompted to identify examples of pain in *The Water Dancer*, Chat GPT identified  
 11 eight examples, including the "physical pain of slavery" and the "pain of longing for freedom"  
 12 and how the characters "yearn for freedom and suffer the pain of not knowing if it will ever be  
 13 attainable." When asked to write a paragraph in the style of *The Water Dancer*, ChatGPT  
 14 generated passages imitating Plaintiff Coates's writing style including references to the  
 15 "defy[ing] the chains of oppression" and the "dreams of people yearning to be free." *Exhibit B*.

16       56. ChatGPT also provided in depth summaries and analyses of Plaintiff Díaz's  
 17 novel *The Brief Wondrous Life of Oscar Wao*. For instance, when prompted to provide examples  
 18 of the immigrant experience in *The Brief Wondrous Life of Oscar Wao*, ChatGPT identified  
 19 eight specific examples, including cultural traditions, "such as the belief in the 'fukú' curse and  
 20 the significance of family and community" and how "cultural elements serve as a link to their  
 21 roots and provide a sense of continuity in their lives." When asked to write a paragraph in the  
 22 style of *The Brief Wondrous Life of Oscar Wao*, Chat GPT generated passages imitating  
 23 Plaintiff Díaz's style including references to the "the 'fukú'curse" and "what it meant to be both  
 24 Dominican and American." *Exhibit C*.

25       57. When prompted, ChatGPT accurately summarized Plaintiff Greer's novel, *The*  
 26 *Confessions of Max Tivoli*. When prompted to identify themes of the nature of beauty in *The*  
 27 *Confessions of Max Tivoli*, ChatGPT identified six specific themes of beauty in the novel,  
 28 including "ephemeral beauty." When prompted to write a paragraph in the style of *The*

1     *Confessions of Max Tivoli*, ChatGPT generated passages imitating Plaintiff Greer's style,  
 2 including references to beauty as a "shifting, mutable canvas." *Exhibit D*.

3         58.     ChatGPT similarly provided in depth summaries and analyses of Plaintiff  
 4 Hwang's play, *The Dance and the Railroad*. For example, when prompted, ChatGPT identified  
 5 five key themes from *The Dance and the Railroad*, including "art and creativity as a form of  
 6 resistance" and "using art as a form of escape from the harsh realities and dehumanization of  
 7 labor." Additionally, when prompted to produce a screenplay in the style of *The Dance and the*  
 8 *Railroad*, ChatGPT produced a script written in Plaintiff Hwang's style, which generated a  
 9 screenplay involving a Chinese laborer toiling on the Central Pacific Railroad that "believe[s]  
 10 in the power of art to keep [their] spirits alive." *Exhibit E*.

11         59.     Likewise, ChatGPT provided in depth summaries and analyses of Plaintiff  
 12 Klam's works. For example, when prompted, Chat GPT accurately summarized Plaintiff Klam's  
 13 novel *Who is Rich?* and correctly analyzed the key relationships between the novel's central  
 14 character and the other characters in the novel. When asked to identify the main themes in *Who*  
 15 *is Rich?* Chat GPT accurately identified seven main themes of the novel including "mid-life  
 16 crisis and identify." Further, when prompted to write a paragraph in the style of *Who is Rich?*,  
 17 ChatGPT generated passages imitating Plaintiff Klam's writing style, including a reference to  
 18 navigating the "treacherous waters of midlife." *Exhibit F*.

19         60.     When prompted, ChatGPT accurately summarized Plaintiff Lippman's mystery  
 20 novel *What the Dead Know*. When prompted to identify the main themes in *What the Dead*  
 21 *Know*, ChatGPT identified seven main themes of the novel including "identity and memory."  
 22 Further, when prompted to write a paragraph in the style of *What the Dead Know*, ChatGPT  
 23 generated passages imitating Plaintiff Lippman's writing style, including references to the  
 24 "recesses of memory's labyrinth." *Exhibit G*.

25         61.     In the same vein, after being prompted to summarize Plaintiff Snyder's book,  
 26 *What We've Lost is Nothing*, ChatGPT accurately identified themes included within the novel,  
 27 such as "safety, perception, and the fragility of human relationships." Similarly, once prompted,  
 28 ChatGPT accurately analyzed the theme of safety using a specific example from the text of

1 Plaintiff Snyder's copyrighted work, explaining that "the theme of safety is examined through  
 2 the lens of a series of burglaries that occur in a suburban neighborhood . . . and how these  
 3 incidents affect the characters and their perceptions of the world around them." ChatGPT was  
 4 also able to generate random passages authentically written in Plaintiff Snyder's writing style  
 5 when prompted. *Exhibit H.*

6 62. Additionally, ChatGPT provided in depth summaries and analyses of Plaintiff  
 7 Waldman's works. For instance, when prompted to summarize Plaintiff Waldman's novel *Love*  
 8 and *Other Impossible Pursuits*, Chat GPT accurately provided a summary and analysis of the  
 9 novel. When prompted to identify specific instances of grief in *Love and other Impossible*  
 10 *Pursuits*, ChatGPT identified five specific instances of grief, including the protagonist Emelia's  
 11 loss of her infant daughter, a "loss that occurred before the events of the novel and [that] continue  
 12 to haunt Emelia, affecting her emotional state and relationships." When prompted to write a  
 13 paragraph in the style of *Love and Other Impossible Pursuits*, ChatGPT generated a paragraph  
 14 imitating Plaintiff Waldman's writing style, including references to the "weight of her  
 15 daughter's absence." *Exhibit I.*

16 63. Similarly, when prompted, Chat GPT accurately summarized Plaintiff  
 17 Woodson's novel *Brown Girl Dreaming*. For example, when prompted to provide the main  
 18 themes in *Brown Girl Dreaming*, ChatGPT identified seven specific themes, including "racial  
 19 injustice and civil rights" and "family and belonging." When prompted to write a paragraph in  
 20 the style of *Brown Girl Dreaming* ChatGPT generated passages imitating Plaintiff Woodson's  
 21 style, including references that the "struggle for justice and equality colored the air we breathed"  
 22 and the "comforting embrace of family." *Exhibit J.*

23 64. At no point did ChatGPT reproduce any of the copyright management  
 24 information Plaintiffs included with their published works.

25 65. Furthermore, at no point did Plaintiffs authorize OpenAI to download and copy  
 26 their protected works, as described above.

27 **CLASS ALLEGATIONS**

28 66. Plaintiffs bring this action pursuant to the provisions of Rules 23(a), 23(b)(2),

1 and 23(b)(3) of the Federal Rules of Civil Procedure, on behalf of themselves and the following  
 2 proposed Class:

3 All persons or entities in the United States that own a United States copyright in  
 4 any written work that OpenAI used to train any GPT model during the Class  
 Period.

5 67. Excluded from the Class are Defendant, its employees, officers, directors, legal  
 6 representatives, heirs, successors, wholly- or partly-owned, and its subsidiaries and affiliates;  
 7 proposed Class counsel and their employees; the judicial officers and associated court staff  
 8 assigned to this case and their immediate family members; all persons who make a timely  
 9 election to be excluded from the Class; governmental entities; and the judge to whom this case  
 10 is assigned and his/her immediate family.

11 68. This action has been brought and may be properly maintained on behalf of the  
 12 Class proposed herein under Federal Rule of Civil Procedure 23.

13 69. Numerosity. Federal Rule of Civil Procedure 23(a)(1): The members of the Class  
 14 are so numerous and geographically dispersed that individual joinder of all Class members is  
 15 impracticable. On information and belief, there are at least tens of thousands of members in the  
 16 Class. The Class members may be easily derived from Defendants' records.

17 70. Commonality and Predominance. Federal Rule of Civil Procedure 23(a)(2) and  
 18 23(b)(3): This action involves common questions of law and fact, which predominate over any  
 19 questions affecting individual Class members, including, without limitation:

- 21 a. Whether Defendants engaged in the conduct alleged herein;
- 22 b. Whether Defendants violated the copyrights of Plaintiffs and the Class when they  
     downloaded and copied Plaintiffs' and the Class's copyrighted books;
- 23 c. Whether ChatGPT itself is an infringing derivative work based on Plaintiffs' and  
     the Class's copyrighted books;
- 24 d. Whether the text responses of ChatGPT are infringing derivative works based on  
     Plaintiffs' and the Class's copyrighted books;
- 25 e. Whether Defendants violated the DMCA by removing copyright-management

1 information from Plaintiffs' and the Class's copyrighted books;

2 f. Whether Defendants were unjustly enriched by the unlawful conduct alleged  
3 herein;

4 g. Whether Defendants' conduct violates the California Unfair Competition Law;

5 h. Whether Plaintiffs and the other Class members are entitled to equitable relief,  
6 including, but not limited to, restitution or injunctive relief; and

7 i. Whether Plaintiffs and the other Class members are entitled to damages and other  
8 monetary relief and, if so, in what amount.

9 71. Typicality. Federal Rule of Civil Procedure 23(a)(3): Plaintiffs' claims are  
10 typical of the other Class members' claims because, among other things, all Class members were  
11 comparably injured through Defendants' wrongful conduct as described above.

12 72. Adequacy. Federal Rule of Civil Procedure 23(a)(4): Plaintiffs are adequate  
13 Class representative because their interests do not conflict with the interests of the other  
14 members of the Class they seeks to represent; Plaintiff have retained counsel competent and  
15 experienced in complex class action litigation; and Plaintiffs intend to prosecute this action  
16 vigorously. The interests of the Class will be fairly and adequately protected by Plaintiffs and  
17 their counsel.

18 73. Declaratory and Injunctive Relief. Federal Rule of Civil Procedure 23(b)(2):  
19 Defendants have acted or refused to act on grounds generally applicable to Plaintiffs and the  
20 other members of the Class, thereby making appropriate final injunctive relief and declaratory  
21 relief with respect to the Class as a whole.

22 74. Superiority. Federal Rule of Civil Procedure 23(b)(3): A class action is superior  
23 to any other available means for the fair and efficient adjudication of this controversy, and no  
24 unusual difficulties are likely to be encountered in the management of this class action. The  
25 damages or other financial detriment suffered by Plaintiffs and the other Class members are  
26 relatively small compared to the burden and expense that would be required to individually  
27 litigate their claims against Defendants, so it would be impracticable for the members of the  
28 Class to individually seek redress for Defendants' wrongful conduct. Even if Class members

1 could afford individual litigation, the court system could not. Individualized litigation creates a  
 2 potential for inconsistent or contradictory judgments, and increases the delay and expense to all  
 3 parties and the court system. By contrast, the class action device presents far fewer management  
 4 difficulties, and provides the benefits of single adjudication, economy of scale, and  
 5 comprehensive supervision by a single court.

6 **CAUSES OF ACTION**

7 **FIRST CAUSE OF ACTION**

8 **DIRECT COPYRIGHT INFRINGEMENT,**  
 9 **17 U.S.C. § 106, *et seq.***

10 75. Plaintiffs hereby incorporate by reference the allegations contained in the  
 preceding paragraphs of this Complaint.

11 76. Plaintiffs bring this claim on behalf of themselves and on behalf of the Class  
 against Defendants.

12 77. As the owners of the registered copyrights in books used to train OpenAI's GPT  
 models, Plaintiffs and the Class hold the exclusive rights to those works under 17 U.S.C. § 106.

13 78. Plaintiffs have obtained copyright registrations for each of the works identified  
 in *Exhibit K*.

14 79. On information and belief, to train OpenAI's GPT models, OpenAI relied on  
 harvesting mass quantities of content from the public internet, including Plaintiffs' and the  
 Class's written works, which are available in digital formats.

15 80. Because OpenAI's GPT models cannot function without the expressive  
 information extracted from Plaintiffs' and Class members' works and retained by the GPT  
 models, GPT and ChatGPT are themselves infringing derivative works, made without Plaintiffs'  
 and Class members' permission in violation of their exclusive rights under the Copyright Act.

16 81. Plaintiffs and the Class never authorized OpenAI to make copies of their written  
 works, make derivative works, publicly display copies (or derivative works), or distribute copies  
 (or derivative works). Each of those rights belong exclusively to Plaintiffs and Class members  
 under copyright law.

82. By and through the actions alleged above, OpenAI has infringed and will continue to infringe Plaintiffs' and the Class's copyrights.

83. OpenAI's acts of copyright infringement have been intentional, willful, and in callous disregard of Plaintiffs' and Class members' rights. OpenAI knew at all relevant times that the datasets it used to train its GPT models contained copyrighted materials, and that its acts were in violation of the terms of use of the materials.

84. OpenAI engaged in the infringing acts described herein for its own commercial benefit.

85. As a direct and proximate result of OpenAI's wrongful conduct, Plaintiffs have been substantially and irreparably injured by OpenAI's acts of direct copyright infringement in an amount not readily capable of determination and, unless permanently enjoined from further acts of infringement and continuing to use and distribute GPT models trained using Plaintiffs' and Class members' copyrighted materials without permission, OpenAI will cause additional irreparable harm for which there is no adequate remedy at law. Plaintiff and the Class are thus entitled to permanent injunctive relief preventing OpenAI from engaging in any further infringement of Plaintiffs' and the Class's copyrighted works.

86. Plaintiffs are further entitled to recover statutory damages, actual damages, restitution of profits, and other remedies provided by law.

## **SECOND CAUSE OF ACTION**

**VICARIOUS COPYRIGHT INFRINGEMENT  
17 U.S.C. § 106**

87. Plaintiffs incorporate by reference all allegations of the preceding paragraphs as though fully set forth herein.

88. Plaintiffs bring this claim on behalf of themselves and on behalf of the Class against Defendants.

89. Defendant OpenAI, LP is the for-profit subsidiary of Defendant OpenAI, Inc. and is principally responsible for and dedicated to the development of the GPT models and ChatGPT products at issue in this action. Defendant OpenAI Startup Fund Management, LLC

1       exercised control over Defendant OpenAI, LP, along with Defendant OpenAI GP, LLC, which  
 2       is the general partner of Defendant OpenAI, LP, responsible for managing and operating the  
 3       day-to-day business affairs of Defendant OpenAI, LP, and is wholly owned and controlled by  
 4       Defendant OpenAI, Inc., along with Defendant OpenAI OpCo, LLC. Upon information and  
 5       belief, Defendant OpenAI Startup Fund I, LP played a vital role in the foundation of Defendant  
 6       OpenAI, LP, including providing initial funding and creating its business strategy, while  
 7       Defendant OpenAI Startup Fund GP I, LLC is the general partner of Defendant OpenAI Startup  
 8       Fund I, LP, responsible for managing and operating the day-to-day business affairs of Defendant  
 9       OpenAI Startup Fund I, LP.

10       90.       Defendant OpenAI, LP directly infringed upon Plaintiffs' and Class members'  
 11       copyrighted works through the unauthorized use and reproduction of the works, and preparation  
 12       of derivative works by ChatGPT. As discussed above, Plaintiffs' and Class members' protected  
 13       works were used to train GPT models. Because the GPT models are based on expressive  
 14       information extracted from Plaintiffs' and Class members' works, Defendant OpenAI, LP is  
 15       directly liable for unauthorized use, reproduction, display of copyrighted works, as well as  
 16       creation of derivative works through ChatGPT's responses. Therefore, Defendant OpenAI, LP  
 17       directly infringed upon Plaintiffs' and Class members' exclusive rights under 17 U.S.C. § 106.  
 18

19       91.       Defendants OpenAI, Inc., OpenAI OpCo, LLC, OpenAI GP, LLC, OpenAI  
 20       Startup Fund GP I, LLC, OpenAI Startup Fund I, LP, and OpenAI Startup Management LLC  
 21       are vicariously liable for the infringement alleged herein because they had the right and ability  
 22       to supervise and control the infringing activity but failed to stop the infringing conduct.  
 23

24       92.       Furthermore, Defendants have a direct financial interest in the infringing conduct  
 25       and received revenue in connection with the development, deployment, and advancement of the  
 26       GPT models and ChatGPT. Each entity profited from the advancement of GPT models and  
 27       ChatGPT.  
 28

93. These committed acts of copyright infringement were willful, intentional, and malicious and thus subjects Defendants to liability for statutory damages under Section 504(c)(2) of the Copyright Act of up to \$150,000 per infringement.

94. Plaintiffs and Class members have been injured by Defendants' acts of vicarious copyright infringement. Plaintiffs and the Class are entitled to statutory damages, actual damages, restitution of profits, and other remedies provided by law.

### **THIRD CAUSE OF ACTION**

**DIGITAL MILLENNIUM COPYRIGHT ACT – REMOVAL OF COPYRIGHT  
MANAGEMENT INFORMATION  
17 U.S.C. § 1202(B)**

95. Plaintiffs incorporate by reference all allegations of the preceding paragraphs as though fully set forth herein.

96. Plaintiffs bring this claim on behalf of themselves and on behalf of the Class against Defendants.

97. Plaintiffs and Class members included one or more forms of copyright-management information in each of Plaintiffs' and Class members' infringed works, including: copyright notice, title and other identifying information, the name or other identifying information about the owners of each book, terms and conditions of use, and identifying numbers or symbols referring to the copyright-management information.

98. Without the authority of Plaintiffs and the Class, OpenAI copied Plaintiffs' and Class members' works and used them as training data for its GPT software. By design, the training process does not preserve any copyright-management information. Therefore, OpenAI intentionally removed copyright-management information from Plaintiffs' and Class members' works in violation of 17 U.S.C. § 1202(b)(1).

99. OpenAI's removal or alteration of Plaintiffs' and the Class's copyright-management information has been done knowingly and with the intent to induce, enable, facilitate, or conceal infringement of Plaintiffs' and the Class's copyrights.

100. Without the authority of Plaintiffs and the Class, Defendants created derivative

1 works based on Plaintiffs' and Class members' works. By distributing these works without their  
 2 copyright-management information, OpenAI violated 17 U.S.C. § 1202(b)(3).

3 101. OpenAI knew or had reasonable grounds to know that this removal of copyright-  
 4 management information would facilitate copyright infringement by concealing the fact that  
 5 every output from ChatGPT is an infringing derivative work, synthesized entirely from  
 6 expressive information found in the training data.

7 102. Plaintiffs and the Class have been injured by OpenAI's removal of copyright-  
 8 management information. Plaintiffs and the Class are entitled to statutory damages, actual  
 9 damages, restitution of profits, and other remedies provided by law, including full costs and  
 10 attorneys' fees.

11 **FOURTH CAUSE OF ACTION**

12 **VIOLATIONS OF THE CALIFORNIA UNFAIR COMPETITION LAW**  
 13 **CAL. BUS. & PROF. CODE §§ 17200, ET SEQ.**

14 103. Plaintiffs and the Class incorporate by reference each preceding and succeeding  
 15 paragraph as though fully set forth at length herein.

16 104. Plaintiffs bring this claim on behalf of themselves and on behalf of the Class  
 17 against Defendants.

18 105. The California Unfair Competition Law ("UCL") prohibits acts of "unfair  
 19 competition," including any "unlawful, unfair or fraudulent business act or practice" and "unfair,  
 20 deceptive, untrue or misleading advertising." Cal. Bus. & Prof. Code § 17200.

21 106. Defendants have engaged in unfair competition and unfair, unlawful or  
 22 fraudulent business practices by the conduct, statements, and omissions described above because  
 23 it illegally collected and used Plaintiffs' and the Class's copyrighted works to train its GPT  
 24 models.

25 107. The unlawful business practices described herein violate the UCL because  
 26 Defendants used Plaintiffs' and the Class's protected works to train its GPT software for  
 27 Defendants' own commercial profit without Plaintiffs' and the Class's authorization.  
 28 Defendants further knowingly designed ChatGPT to include portions or summaries of Plaintiffs'

copyrighted works without attributions in its responses, and they unfairly profit from and take credit for developing a commercial product based on unattributed reproductions of those stolen writing and ideas.

108. These acts and practices have deceived Plaintiffs and are likely to deceive the public into believing that Plaintiffs and the Class have granted OpenAI the right to use its copyrighted materials. In failing to disclose the sources of its training datasets and suppressing other material facts from Plaintiffs and Class members as well as the public, Defendant breached its duties to disclose these facts, violated the UCL, and caused injuries to Plaintiffs and Class members. The omissions and acts of concealment by Defendants pertained to information that was material to Plaintiffs and Class members, as it would have been to all reasonable consumers.

109. The injuries suffered by Plaintiffs and Class members are not greatly outweighed by any potential countervailing benefit to consumers or to competition, nor are they injuries that Plaintiffs and Class members should have reasonably avoided.

110. Defendant's acts and practices are unlawful because they violate California Civil Code §§ 1668, 1709, 1710, and 1750 *et seq.*, and California Commercial Code § 2313.

111. Plaintiffs seek to enjoin further unlawful, unfair and/or fraudulent acts or practices by Defendants, to obtain restitutionary disgorgement of all monies and revenues generated as a result of such practices, and all other relief allowed under California Business & Professions Code § 17200.

## **FIFTH CAUSE OF ACTION**

## NEGLIGENCE

112. Plaintiffs incorporate by reference the allegations of all foregoing paragraphs as if they had been set forth in full herein.

113. Plaintiffs bring this claim on behalf of themselves and on behalf of the Class against Defendants.

114. Defendants owed a duty of care toward Plaintiffs and the Class in (1) obtaining data to train its GPT models and (2) not using Plaintiffs' and the Class's protected works to train its GPT models.

115. Defendants have a common law duty to prevent foreseeable harm to others, including Plaintiffs and members of the Class, who were foreseeable and probable victims of Defendants' unlawful practices.

116. Defendants breached their duty to exercise due care by negligently, carelessly, and recklessly collecting, maintaining, and controlling Plaintiffs' and Class members' works and engineering, designing, maintaining, and controlling systems—including ChatGPT—that are trained on Plaintiffs' and Class members' works without their authorization.

117. The damages suffered by Plaintiffs and the Class were the direct and reasonably foreseeable result of Defendants' negligent breach of their duties to adequately design, implement, and maintain reasonable practices to avoid infringing protected works without consent of copyright holders.

118. Defendants' negligence directly caused significant harm to Plaintiffs and the Class.

## **SIXTH CAUSE OF ACTION**

## UNJUST ENRICHMENT

119. Plaintiffs incorporate by reference all allegations of the preceding paragraphs as though fully set forth herein.

120. By virtue of the unlawful, unfair, and deceptive conduct alleged herein, Defendants knowingly realized substantial revenue from the use of Plaintiffs' and Class members' copyrighted works for the commercial training of its GPT models used to power its ChatGPT product.

121. Defendant knew or should have known that Plaintiffs and the Class have invested substantial time and energy creating the works in which they hold a copyright.

122. Defendants were conferred significant benefits when they downloaded and copied Plaintiffs' and the Class's copyrighted works to train their GPT software without Plaintiffs' and the Class's permission. Defendant knowingly and willingly accepted and enjoyed those benefits.

123. By using Plaintiffs' and the Class's copyrighted works to train ChatGPT,

Defendants caused Plaintiffs and the Class to suffer actual damages from the deprivation of the benefits of their work, including monetary damages.

124. Defendants derived profit and other economic benefits from the use of Plaintiffs' and the Class's copyrighted works to train ChatGPT.

125. It would be inequitable and unjust to permit Defendants to retain the enormous economic benefits it has obtained from and/or at the expense of Plaintiffs and Class members.

126. As a direct and proximate cause of Defendants' unjust enrichment, Plaintiffs and the Class are entitled to restitution, attorneys' fees, costs and interest.

Defendants' conduct is causing and, unless enjoined and restrained by this Court, will continue to cause Plaintiffs and the Class irreparable injury that cannot be compensated or measured in money.

## **REQUEST FOR RELIEF**

WHEREFORE, Plaintiffs, individually and on behalf of members of the Class defined above, respectfully request that the Court enter judgment against Defendants and award the following relief:

A. Certification of this action as a class action pursuant to Rule 23 of the Federal Rules of Civil Procedure, declaring Plaintiffs as the representative of the Class, and Plaintiffs' counsel as counsel for the Class;

B. An order awarding declaratory relief and temporarily and permanently enjoining Defendant from continuing the unlawful and unfair business practices alleged in this Complaint and to ensure that all applicable information set forth in 17 U.S.C. § 1203(b)(1) is included when appropriate;

C. An award of statutory and other damages under 17 U.S.C. § 504 for violations of the copyrights of Plaintiff and the Class by Defendants.

D. An award of statutory damages under 17 U.S.C. § 1203(b)(3) and 17 U.S.C. § 1203(c)(3), or in the alternative, an award of actual damages and any additional profits under 17 U.S.C. § 1203(c)(2);

E. A declaration that Defendant is financially responsible for all Class notice and

1 the administration of Class relief;

2 F. An order awarding any applicable statutory and civil penalties;

3 G. An order requiring Defendant to pay both pre- and post-judgment interest on any  
4 amounts awarded;

5 H. An award of costs, expenses, and attorneys' fees as permitted by law; and

6 I. Such other or further relief as the Court may deem appropriate, just, and  
7 equitable.

8 **DEMAND FOR JURY TRIAL**

9 Plaintiffs hereby demand a jury trial for all claims so triable.

10  
11 DATED: October 5, 2023

Respectfully submitted,

12 /s/ *Daniel J. Muller*

13 DANIEL J. MULLER, SBN 193396  
14 dmuller@venturahersey.com  
15 **VENTURA HERSEY & MULLER, LLP**  
16 1506 Hamilton Avenue  
17 San Jose, California 95125  
Telephone: (408) 512-3022  
Facsimile: (408) 512-3023  
dmuller@venturahersey.com

18 /s/ *Bryan L. Clobes*

19 Bryan L. Clobes (*pro hac vice anticipated*)  
20 **CAFFERTY CLOBES MERIWETHER**  
21 & SPRENGEL LLP  
22 205 N. Monroe Street  
Media, PA 19063  
Tel: 215-864-2800  
bclobes@caffertyclobes.com

23 Alexander J. Sweatman (*pro hac vice anticipated*)  
24 **CAFFERTY CLOBES MERIWETHER**  
25 & SPRENGEL LLP  
26 135 South LaSalle Street, Suite 3210  
27 Chicago, IL 60603  
Tel: 312-782-4880  
asweatman@caffertyclobes.com

28 *Attorneys for Plaintiffs*